IEEE CIS

AMD NEWSLETTER

The Newsletter of the Autonomous Mental Development Technical Committee

Developmental Robotics
Machine Intelligence
Neuroscience
Psychology

Volume 12, number 1
Spring 2015

Editorial: Building Up the Community: Interdisciplinary Bridges and Open Science

Pierre-Yves Oudeyer

Inria and Ensta
ParisTech, France
Editor & AMD TC Chair
pierre-yves.oudeyer@inria.fr

Dialogues

The scientific dialogue in this newsletter's issue, proposed by Janet Wiles, revolves around the question "Will social robots need to be consciously aware?". Responses are provided by Axel Cleeremans, Yasuko Kitano, Cornelius Weber and Stefan Wermter, Justin Hart and Brian Scassellati, Juyang Weng, Guy Hoffman and Moran Cerf. Several dimensions of the question stand out. First, as we are very far from understanding what "consciousness" is, it appears that building robots capable of various forms of self- and other- awareness, and importantly how they can develop these capabilities progressively, can be very useful in the quest to unveil the underlying mechanisms. Second, as consciousness is a multiscale complex systems, multiple approaches and perspectives need to be taken in this process of robot building. Third, when one looks at applications, it is the function, and not the nature, of consciousness which becomes the relevant angle of analysis, and several ethical questions arise.

Then, a new dialog is initiated by Stéphane Doncieux on the topic of representational redescription. It has long been known in AI that having a good representation is key for machines to solve complex problems. However, so far good representations have been pre-programmed by engineers. What technical approaches could we imagine to allow machines to select, and even more important to find, new spaces of representations? Are techniques like deep learning general enough for realising such a challenge for life-long learning robots? Do we need other approaches such as Darwinian mechanisms operating in the brain, like in neural Darwinism? Those of you interested in reacting to this dialogue initiation are welcome to submit a response by October 30th, 2015. The length of each response must be between 600 and 800 words including references (contact pierre-yves.oudeyer@inria.fr).

New AMD TC Chair

I am writing this editorial not only as editor of the newsletter, but also as the new Chair of the IEEE CIS Technical Committee on Autonomous Mental Development. I am very honoured to be appointed to this job, and I would first like to thank Matthew Schlesinger and my other predecessors for the amazing work they have been doing to stimulate the community. As we are still a developing community :), there are significant organisational and strategic challenges that still need to be addressed, and to which I will do my best to contribute.

First, and due to the fundamental interdisciplinary character of our field, there are many research communities who have been working on topics related to computational modelling of development in machines and animals, but who have been striving in a relative isolation. We need to build bridges and connections with research communities such as connections modelling, cognitive systems and AI, evolutionary systems, psychology and neuroscience.

As a first action, I have proposed the creation of a new task force of the AMD TC related to a field I believe should become very important in the years to come: Evolutionary-Developmental robotics. This task force's chair and vice-chairs are Jean-Baptiste Mouret, Jeff Clune and Stéphane Doncieux, who are among the most creative researchers in this area. A second action is to strengthen the link with a solid community of researchers who for years have been developing connectionist constructivist models of cognitive development in the connectionist communities, and I am happy

that Gert Westermann and Denis Mareschal, key actors of this trend, are chair and vice-chair of the TF on Developmental Psychology.

As a second action and in this context, the TC continues to contribute to the reflection about the evolution of the scope of the TAMD journal and the TC itself (and to the change of their names), in collaboration with the new editor Angelo Cangelosi, whose great experience will be very helpful to continue to strengthen the journal. An evolution will be to cover topics beyond developmental issues (in practice, the journal has been accepting around 1/3rd of papers not focused on development, but on cognition in general). The idea is to keep development as a very strong component, but contextualized with other work to build connections (See the recent editorial by Angelo in TAMD vol 7. issue 1).

Second, we need flexible tools allowing the community to discuss and share ideas in an open manner. A first step in this direction is the opening of a web space for open discussion, based on the Discourse forum technology, which allows the easy creation of multimedia exchanges, and can be used to foster open science discussions. The web pages of the AMD TC and TFs are now hosted there. Anyone wanting to start a dialog about any topic linked to computational modelling of development, or related to the organisation of the community, can use the tool. Begin the discussion at: www.icdl-epirob.org/amdtc

The community has also now an active Twitter account that everyone shall use to give and read news: @DevRobNews (thanks a lot to Matthias Rolf and Alessandra Sciutti for animating the Web TF!). See also the ICDL-Epirob twitter account: @ICDL_Epirob2015

Call for candidate for becoming the new editor of the newsletter

As I have been editor of the newsletter for around 8 years, and I have now the AMD TC Chair duty, it is now time for other ideas and energies to take the lead of the newsletter. This has been a fantastic job, especially in the organisation of scientific dialogues with many of the major thinkers of our community. If you are interested in the job, please send your application to pierre-yves.oudeyer@inria.fr, which will then be reviewed by the members of the TC.

ICDL-Epirob 2015 and call for organisation of next editions

This summer, the ICDL-Epirob conference is taking place at Brown University, Providence, USA, August 13th-16th, and will feature three outstanding keynote speakers: Dare Baldwin (Univ. Oregon, US), Kerstin Dautenhahn (Univ. Hertfordshire, UK), and Asif Ghazanfar (Princeton Univ., US). This year the general chairs are Dima Amso and Matthew Schlesinger, and the program chairs are Anne Warlaumont and Clément Moulin-Frier, showing a significant emphasis on developmental psychology and neuroscience. This is in practice embodied through a great initiative by Matthew Schlesinger and Anne Warlaumont who have proposed the Babybot Challenge at ICDL-Epirob to strengthen the link between scientists studying development in humans and in artificial systems. The challenge is to select from a list of three infant studies and design a model that captures infants' performance on the chosen task. The results will be announced during the conference.

In 2016, ICDL-Epirob is planned to take place in Paris, France, thanks to the proposal of Philippe Gaussier and his colleagues at University Cergy-Pontoise. For the following years, all applications for organisation are welcome!

Table of Contents

Dialogue

Will Social Robots Need to Be Consciously Aware?

Janet Wiles

Complex & Intelligent Systems Research Group, School of Information Technology & Electrical Engineering,
The University of Queensland
St Lucia, Australia

j.wiles@uq.edu.au

For autonomous robots to take social roles in homes, health, education, or entertainment, they will require a range of cognitive and social abilities. In this dialog, I focus on the different types of awareness required to underpin social interaction. Let us start with a broader version of the question in the title:

What aspects of awareness will an autonomous robot need for interpersonal engagement with humans and other autonomous agents?

Human social skills are deeply rooted in mammalian biology, and interactions between robots and non-human animals can reveal biological bases of social interactions in ways that are not possible or ethical with humans alone. One such robot, the iRat (see Figure, (Ball et al. 2010)) is currently being developed as a social companion for studies in rodent social neuroscience. The iRat is intentionally minimalistic – its form is a simple oval shape with no external limbs or moving parts. Its only behaviour is movement, and to date it has only rudimentary social abilities. However, even such simple abilities are sufficient to engage the interest of real rats (Wiles et al., 2012), and its social successes and failures provide a starting point for discussion.

Embodiment matters for social engagement: For the iRat to become an effective social companion, it needs to move in the same laboratory environments as rats. The iRat's most important physical feature is that it is rat-sized, and can operate safely in close proximity with real rats. Rats will readily explore the iRat when it is stationary and moving, sniffing, touching, whisking, and on some occasions even riding on it.

What does a social robot need to know about its physical and social world?

Awareness of social spatial relationships: The iRat has an awareness of space provided by a bio-inspired navigation system called RatSLAM (Ball et al., 2013; Milford et al. 2010), which mimics place and grid cells (dubbed the brain's "internal GPS system" in a 2014 Nobel award (O'Keefe & Dostrovsky, 1971; Hafting et al., 2005)). But knowing GPS coordinates is not sufficient. The iRat also needs an awareness of spatial relationships with other social beings, including significant social behaviours. A rat that approaches nose to nose with the iRat behaves differently if the iRat retreats, than if it turns aside using an obstacle avoidance behaviour. To understand the meaning of a social spatial relationship requires an

Rat and iRat

awareness of others. In another encounter, a rat approached the iRat from behind and appeared to tap the iRat and then retreat. With a real rat this could have been the prelude to a play sequence. The iRat didn't notice – couldn't notice – because it doesn't yet have a sense of touch to its own body. Social spatial relationships require an awareness of self.

What minimum awareness of self and others is required by a social robot?

Awareness of individual identities, and the interaction histories that go with them: Social engagement is full of episodic encounters. In one environment, a rat was repeatedly visiting a circuit of food chambers, and the iRat was meant to retreat submissively as the rat approached. However, a glitch caused the robot to stall and block the rat's path to its food, a behaviour that could be interpreted as aggressive. On the next two circuits, the rat avoided the iRat and skipped that food chamber completely, forfeiting its reward. Social encounters do not just engage emotional states, or semantic memory. They also create personal histories and episodic memories that are unique to the participants in the encounter. Such episodic memories are the basis of an ability to remember where and when an aggressive (or pleasant) interaction occurred, and with whom, and ultimately, the ability to make and uphold social contracts.

Could a social robot have a subjective world?

Even stronger than "could" I think that social robots will "require" some form of subjective world. It is possible to imagine a social robot without feelings, but that does not make it viable in practice. They might not be the same as yours or mine, but a social robot that has an episodic memory of the events in its past must have a first person experience that is unique to itself. Non-social robots (such as a self-driving car) don't necessarily need one.

Robots with a subjective sense of self open the Pandora's Box of consciousness, a term that is ill-defined for practical robotics. However, recent theories in neuroscience have explored the unity of the conscious self within an overarching framework called Integrated Information Theory (IIT, (Tononi, 2004)). IIT was developed solely from a human first person perspective, but one could imagine extending the ideas (and mathematics) of integrated information to the integration that underpins coherent decisions – the ability to decide; integrated intention – the unity of agency; and integrated perception and action – the unity of a stable sensorimotor experience.

At a recent workshop on Panpsychism – a doctrine that everything has a degree of individual consciousness – a few (3) members of the audience took the position that the iRat is already conscious, albeit at a low level (Tsuchiya et al., 2014). Others argued that robots can never be conscious.

As we design new abilities for future generations of iRats and other social robots, we cannot include every social ability. Which ones are critical? I don't necessarily want to build conscious awareness into a robot, but if the subjective self has a social function, it may be that at least some aspects of conscious awareness will be indispensable in the quest for social robots.

For readers who believe that robots cannot be consciously aware (by some definition of consciousness), the question for this dialog could be rephrased as:

What are the limits to the social abilities of a non-conscious robot?

D. Ball, S. Heath, M. Milford, G. Wyeth, and J. Wiles, "A navigating rat animat," in Proceedings of the 12th International Conference on the Synthesis and Simulation of Living Systems, 2010, pp. 804-811.
J. Wiles, S. Heath, D. Ball, L. Quinn, and A. Chiba, "Rat meets iRat," in Development and Learning and Epigenetic Robotics (ICDL), 2012 IEEE International Conference on, 2012, pp. 1-2.
D. Ball, S. Heath, J. Wiles, G. Wyeth, P. Corke, and M. Milford, "OpenRatSLAM: an open source brain-based SLAM system," Autonomous Robots, vol. 34, pp. 149-176, 2013.
M. J. Milford, J. Wiles, and G. F. Wyeth, "Solving navigational uncertainty using grid cells on robots," PLoS computational biology, vol. 6, p. e1000995, 2010.
J. O'Keefe and J. Dostrovsky, "The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat," Brain Research, vol. 34, pp. 171-175, 1971.
T. Hafting, M. Fyhn, S. Molden, M.-B. Moser, and E. I. Moser, "Microstructure of a spatial map in the entorhinal cortex," Nature, vol. 436, pp. 801-806, 2005.
G. Tononi, "An information integration theory of consciousness," BMC neuroscience, vol. 5, p. 42, 2004.
N. Tsuchiya, B. van Swinderen, and O. Carter, Consciousness here, there, everywhere? The prospects for panpsychism. Byron Bay, 2014.

For Whom Robots Are Conscious?

Yasuko Kitano

Department of History and Philosophy of Science,
University of Tokyo,
Japan

yskitano@gmail.com

To answer this question, I would like to introduce a distinction between (1) human-perceived consciousness and (2) consciousness of robots themselves. Regarding (1), the social brain of humans will do its job so that anthropomorphism works in social interaction with robots. Takahashi et al. (2014) suggests that the social machinery in human brains (intuitively) works when we play a simple penny-matching game, not only when playing with a human-like android but also when playing with a non-human-like computer.

To contrast (1) and (2), imagine three types of human-like androids, each of which is made for a different type of social interaction:

(Robot A) for non-verbal interaction: gaze following, pointing, eye contact, and joint attention. Klin et al. (2002) reports that individuals with autism failed to follow the pointing gesture of a character in a movie. Their shifting attention was too delayed to identify the designated target. Robot A has no such difficulty.

(Robot B) for small talk in face-to-face oral conversation: Robot B is still too undeveloped to pass the Turing test and has imperfect joint attention. Yet it is able to differentiate various emotional cues, and its speech can be emotionally nuanced.

(Robot C) for describing the experience of consciousness in temporally distant verbal communications (written language): It can create subjective reports about its own experiences like "This is what it is like to see genuine ultramarine blue, which I have never seen before!" and send back, within the scope of this topic, fairly good responses to yours. Joint attention is irrelevant here because Robot C is designed for written communications.

Humans may perceive the three as mind-holders in an anthropomorphic sense. Although consciousness is to be distinguished from mind in sciences and philosophy, there will be no difference in humans' intuitive reactions. So, in the sense of (1), there seems to be no difference among A, B, and C.

However, in (2), there is. In the philosophical understanding of consciousness, Robot C is

the most promising candidate for conscious robots. In philosophy, conscious experience has traditionally been characterized by quality (qualia) and "aboutness" (intentionality): consciousness is always about something specific such as the redness of red. "Aboutness" is occasionally interpreted as including reflectivity: reflection of one's own experience. Reportability, therefore, has been positioned almost as an a priori of philosophical investigation of consciousness.

Some empirical theories share this philosophical conception. One example is Attention Schema Theory (AST: Graziano, 2013). According to AST, awareness is a model of attention. By providing a sketch of a schema of attention, awareness allows humans to produce and share self-narratives about how the world is framed from their perspective. Since attention has usefulness in predicting behavior (of oneself and others), the emergence of this social device would have been of great significance to their collective survival.

However, non-reflective interpretation of "for themselves" is also possible. Recently in neuroscience, the relationship between awareness, attention, and other meta-cognitive issues has been hotly debated. For example, Wilke et al. (2009) suggests that the neural correlates of qualia can and should be distinguished from neural activity that is associated with cognitive access and perceptual reports.

In summary, "consciously" can be interpreted as describing (1) human-perceived consciousness and (2) consciousness of robots themselves; (2) could be split into two: consciousness of robots themselves in the reflective and non-reflective sense.

Graziano, M.S.A. (2013). Consciousness and the Social Brain (Oxford: OUP).
Klin A, Jones W, Schultz R, Volkmar F, and Cohen D. (2002). Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. Arch. Gen. Psychiatry. 59(9), 809-16.
Takahashi, H., Terada, K. Morita, T. Suzuki, S., Haji, T., Kozima, H., Yoshikawa, M., Matsumoto, Y., Omori, T., Asada, A., and Naito, E. (2014). Different impressions of other agents obtained through social interaction uniquely modulate dorsal and ventral pathway activities in the social human brain. Cortex. 58, 289-300.
Wilke, M., Mueller, K.M., and Leopold, D.A. (2009) Neural activity in the visual thalamus reflects perceptual suppression. Proc. Natl. Acad. Sci. USA 106, 9465-9470.

No Arguments Against Consciously Aware Social Robots

Cornelius Weber
weber@informatik.uni-hamburg.de

Stefan Wermter
wermter@informatik.uni-hamburg.de

Dept. of Computer Science,
University of Hamburg,
Germany

A typical service robot can exhibit some cognitive behaviour by perceiving the environment, by knowing its location and the location of persons, for example, and by using its internal representation of this information to plan a path in this environment (Yan et al. 2013). However, social interaction requires further capabilities. For example, before starting to communicate with a person, a socially behaving robot should first assess whether it is in the person's field of view and whether the person is ready to communicate. A socially-assistive robot with such interaction capabilities has been developed in the KSERA project (Johnson et al. 2014) where the robot considers the user's head pose and gaze detection before engaging in a dialogue, and assesses whether the user shares the attention on an object.

Biological inspiration, like the realization of behavior with artificial neural networks, can guide the design and implementation of cognitive systems. For example, mirror neurons in the brain fire when a person or animal performs a specific action, as well as when one observes another person performing that action. Some learning neural architectures developed in the MirrorBot project (Wermter et al. 2005) represented actions similarly when performed or when perceived. This provides a basis for common referencing, mind-reading, learning from imitation and language.

Verbal communication between a robot and a user is an engineering target, since the user can employ well-tried strategies of human communication. To interact naturally, a robot should know about a person's emotional state and intentions and display traits of consciousness. However, such natural conversation may raise the expectation that the robot's responses have the quality of humans' responses. Limitations of the robot's cognitive capabilities, however, may lead to unsuccessful interactions, disappointing the human conversation partner.

There exists already a robot that promises to rarely disappoint some people: Paro, a robotic version of a baby seal. It does not speak nor do service, but its AI capabilities address the relationship part of care-giving. Interaction with Paro may raise nurturing behaviour, lower blood pressure and reduce depression. However, in some care centres, some people with dementia think it is real (Johnston 2014).

For a more sophisticated robot, it is harder to always match a user's expectations. Autonomous robots of today have partially impressive capabilities, but under tight constraints that reduce reliability in unstructured environments. This is mostly due to limited

sensors and limited cognitive abilities rather than limitations of actuators. A robot's current capabilities in a given situation are hard to infer for a user that does not know the robot's cognitive state, but a robot can convey its internal state by generating speech or gestures. For example, a robot's voice or body pose might indicate that it is currently not certain about its next actions. An implementation of a fear concept may limit a robot's available behaviour (Navarro et al. 2012), and a visible expression of its emulated fear can make its behaviour understandable. While true emotions and full self-awareness today seem impossible to implement in an electronic device, an algorithmic analogue of such feelings may constrain a robot to behave more safely, meaningfully, socially and interpretably.

Self-awareness requires alertness, structured neural activations, attentional selection, and at least short-term memory. Furthermore, motivation and emotions seem also necessary for understanding self-awareness. Moreover, self-awareness seems to be an integrative neural process among several distant brain regions where neurons coordinate via activity propagation and synchronous spiking (Supp et al. 2011). Despite such insights, subjective feelings remain inaccessible. Whether a robot can be fully self-aware or not is disputable, and if its awareness state remains elusive, it may be irrelevant. What matters is that a companion robot interacts with people in a reliable, safe, and meaningful way. This engineering goal will drive the development of robots' "brains" towards some form of apparent awareness.

D. Johnson, R. Cuijpers, J. Juola, E. Torta, M. Simonov, A. Frisiello, M. Bazzani, W. Yan, C. Weber, S. Wermter, N. Meins, J. Oberzaucher, P. Panek, G. Edelmayer, P. Mayer, and C. Beck. Socially-Assistive Robots: A comprehensive approach to extending independent living. International Journal of Social Robotics, Vol. 6, Issue 2, pp. 195-211, Springer, 2014.
A. Johnston. Robotic seals comfort dementia patients but raise ethical concerns. KALW, December 8, 2014. http://kalw.org/post/robotic-seals-comfort-dementia-patients-raise-ethical-concerns
N. Navarro, R. Lowe, S. Wermter. A neurocomputational amygdala model of auditory fear conditioning: A hybrid system approach. Proc, of IJCNN 2012, pp. 214-221, 2012. IEEE.
G.G. Supp, M. Siegel, J.F. Hipp, A.K. Engel. Cortical hypersynchrony predicts breakdown of sensory processing during loss of consciousness. Current Biology, Vol 21, pp. 1988-1993, 2011.
S. Wermter, C. Weber, M. Elshaw. Associative neural models for biomimetic multi-modal learning in a mirror neuron-based robot. In Cangelosi A., Bugmann G. & Borisyuk R. (Eds.), Modeling Language, Cognition and Action. Singapore: World Scientific. pp. 31-46, 2005.
W. Yan, C. Weber, S. Wermter. Learning indoor robot navigation using visual and sensorimotor map information. Frontiers in Neurorobotics, Vol. 7, Issue 15, pp. 1-14, 10.3389/fnbot.2013.00015, 2013.

Self-Awareness and Social Competencies

Justin W. Hart
justin.hart@mech.ubc.ca

Brian Scassellati
scaz@cs.yale.edu

Self-awareness is not a single monolithic ability, but rather is a term that encompasses a diverse set of capabilities that emerge as part of a developmental progression. Many of these capabilities are precursors to proficient social behavior, from understanding one's body and senses (Rochat, 2001) to reasoning about one's mental states as being different from those of others (Scassellati, 2002). For this reason, the question is not whether social robots must be self-aware; but which social skills designers wish to incorporate into their systems, and what forms of self-awareness may be required by the desired set of social competencies. The self-awareness aspects of these systems may represent new challenges to artificial intelligence, but it is a topic that we as a community are prepared to study in a meaningful way.

In recent work we have focused on the task of robot self-modeling (Hart, 2014). This is the process of enabling a robot to learn about itself through data sampled during operation, and is inspired by processes encompassing the development of self-awareness in children. We have constructed a self-model learning system which enables a robot to estimate a model of its arm kinematics through data sampled by the vision system (Hart, 2014). It then uses its kinematic structure as an invariant against which it mutually refines its kinematic and visual calibrations. This process produces a tightly-calibrated self-model, which the robot is then able to use to reason about its environment. The process by which the robot infers this model is inspired by the process by which children learn about their bodies and senses, through using them in conjunction with each other. This is one of the earliest forms of self-knowledge to form during infancy (Rochat, 2001). The robot is able to use its self-model in order to infer the visual perspective of a mirror, allowing the robot to infer the positions of objects reflected in the mirror using its stereo vision system (Hart 2014). This test approximates tests of the use of mirrors as instruments for spatial reasoning; a capability that emerges prior to mirror self-recognition in infants (Bertenthal and Fisher, 1978), and appears in animals that are incapable of passing self-recognition tasks (Heschi and Burkart 2006).

Explicitly reasoning about the self is an important component of many social interactions. In Hart (2014), we proposed an extension of the model that that would allow the system to

Justin W. Hart

Department
of Mechanical
Engineering,
University of British
Columbia,
Vancouver, Canada

Brian Scassellati

Computer Science,
Cognitive Science,
and Mechanical
Engineering,
Yale University,
New Haven, USA

infer the visual perspective of other agents in the interaction. By incorporating eye tracking into the robot's vision system, the robot could predict what a human interlocutor has and has not seen. Note that this is only possible by reflecting on the robot's visual perspective, and that it differs from the other agent's. This form of visual perspective taking is an important social skill, and bears a relationship to Theory of Mind (Scassellati, 2002); by which an agent is aware of its own mental state and that it differs from the mental states of others. Reasoning about others by reflecting on the self, and vice-versa, are important social skills that play roles in learning, collaboration, and competition. In collaborative tasks, an agent may reason about another agent's plan by reflecting on what they would be attempting to accomplish by the same actions. Learning can be accomplished by relating another agent's actions to the skill that an agent is trying to learn. In competitive settings, reasoning about the mental states and visual perspectives of other agents may be key to a winning strategy.

While there are many examples of human-robot interaction in which the robot has no form of self-awareness; self-awareness will become an important aspect of interactions moving forward. The key to making this work is a rigorous study of self-awareness that is grounded in achievable milestones based on our best up-to-date knowledge from psychology and neuroscience. Our efforts towards emulating these capabilities have been encouraging, and we intend to continue our line of inquiry in directions that lead towards more self-aware social human-robot interaction.

J.W. Hart, Robot Self-Modeling. Ph.D. Dissertation, Yale University Department of Computer Science. November, 2014.
P. Rochat, The Infant's World. Cambridge, Massachusetts and London, England: Harvard University Press, 2001.
B. Scassellati, "Theory of mind for a humanoid robot," Autonomous Robots, vol. 12, pp. 13–24, 2002.
B. I. Bertenthal and K. W. Fischer, "Development of self-recognition in the infant," Developmental Psychology, vol. 14, no. 4, pp. 44–50, 1978.
A. Heschl and J. Burkart, "A new mark test for self-recognition in non-human primates," Primates, vol. 47, no. 3, pp. 187–198, 2006.
G. G. Gallup, "Self-awareness and the emergence of mind in primates," American Journal of Primatology, vol. 2, pp. 237–248, 1982.

The Social Roots of Consciousness

Axel Cleeremans

Consciousness, Cognition
& Computation Group,
Center for Research in
Cognition & Neurosciences,
Université libre de
Bruxelles,
Belgium
axcleer@ulb.ac.be

Robots like the iRat offer a tantalising glimpse into a future where living things continuously interact with artificial intelligence at different levels — from social interactions all the way down to cyborg fusion between man and machine.

Wiles asks different questions. One is: "What abilities are necessary for future robots to be successful in their social interactions with living agents?" A second question, different but related to the first one, is: "Is subjective experience necessary for social interactions"? The third, most fundamental question, is: "Can we build a conscious robot"?

I will address each in turn, beginning with the last one: What would it take to build a conscious robot? At this point, nobody has a clue, essentially because nobody knows how the activity of biological machines such as the human brain is associated with subjective experience. All major authors (I should say: all authors) in this domain have their own theory of consciousness, and while there is moderate consensus around specific ideas (i.e., the idea that consciousness depends on a global neuronal workspace that links and amplifies content, the idea that it involves a system's capacity to redescribe its own activity, or the idea that it depends on specific types of functional connectivity in the brain), no theory can claim to offer a convincing account of the mechanisms through which a system —any system— becomes a conscious agent. This, of course, does not mean that it is in principle impossible to build a conscious robot; in fact, there is every reason to believe that it will be possible, at least in the sense that there is no principled fundamental argument against it as long as one agrees that the mind has an objective physical ontology. There is nothing magical about consciousness. Brains are biological machines and consciousness depends on the activity of the brain. Hence, once we understand exactly how brains work, we will understand consciousness.

However, I feel there is an essential aspect missing from most attempts at developing an understanding of consciousness, and it is one where robotics may be particularly helpful: the fact that conscious experiences, by necessity, belong to and are embedded in an agent that can act upon the world, that is, a system that has wants, desires, fears and regrets, as well as an ability to represent itself as having such states. Agenthood requires many features that are simply not considered in current research about the differences between conscious and unconscious processing, such as intentions. Contemporary robots,

in this respect, seem profoundly depressed or atonic: They don't want anything of their own and are more or less incapable of learning what to do in order to fulfil their goals. These are complex issues that are ultimately connected to life and death: Even elementary living organisms strive to survive, or at least to achieve some form of homeostasis. Since robots are immortal by design, however, it is somewhat bewildering how such considerations will apply in the future.

Second, is subjective experience necessary for social interactions? The answer to this question depends on exactly what we mean by social interactions. Many of the interactions we routinely engage in are scripted, and it is thus possible to automatise such interactions fairly easily, as systems like Apple's Siri readily demonstrate. But does this count as a "social" interaction? Not really, that is, no more than the automated interactions we engage in with other people. Genuine social interactions, though, that is, unscripted social interactions, require the ability to represent the mental states (and in particular, their intentions) of other agents in a flexible manner. And this process is greatly enhanced by subjective experience, which makes it possible for me to feel, (veridically or not), the state of mind of someone else. Representing the mental states of other agents is difficult because my mental states are not available to inspection by other agents (or even by myself, as my own ability to discern my own mental states by introspection can itself be demonstrated to be rather limited). Thus, when trying to figure out what someone else is thinking, we are stuck with asking them (which additionally requires trusting that they answer truthfully) or with making assumptions based on observing their behaviour (which additionally requires an ability to understand how overt behaviour relates to intentions).

How does this ability to represent the mental states of other agents get going? While there is considerable debate about this issue, it is probably fair to say that one crucial mechanism involves learning about the consequences of the actions that one directs towards other agents. In this respect, interactions with the natural world are fundamentally different from interactions with other agents, precisely because other agents are endowed with unobservable internal states. If I let a spoon drop on a hard floor, the sound that results will always be the same, within certain parameters that only vary in a limited range. The consequences of my action are thus more or less entirely predictable. But if I smile to someone, the consequences that may result are many. Perhaps the person will smile back to me, but it may also be the case that the person will ignore me or that she will display puzzlement, or even that she will be angry at me. It all depends on the context and on the unobservable mental states that the person currently entertains. Of course, there is a lot I can learn about the space of possible responses based on my knowledge of the person, my history of prior interactions with her, and on the context in which my interactions take place. But the point is simply to say that in order to successfully predict the consequences of the actions that I direct towards other agents, I have to build a model of how these agents work. And this is complex because, unlike what is the case for interactions with the natural world, it is an inverse problem: The same action may result in many different reactions, and those different reactions can themselves be caused by many different internal states.

Based on these observations, one provocative claim about the relationships between self-awareness and one's ability to represent the mental states of other agents ("theory of mind", as it is called) is thus that theory of mind comes first, as the philosopher Peter Carruthers had defended. That is, it is in virtue of my learning to correctly anticipate the consequences of the actions that I direct towards other agents that I end up developing models of the internal states of such agents, and it is in virtue of the existence of such models that I become able to gain insight about myself (more specifically: about my self). Thus, by this view, self-awareness, and perhaps subjective experience itself, is a consequence of theory of mind as it develops over extended periods of social intercourse.

This possibility offers an answer to Wiles's first question: "What abilities are necessary for future robots to be successful in their social interactions with living agents?" I would surmise that three features are necessary in this respect: (1) Massive information-processing resources that are sufficiently powerful to simulate certain aspects of their own physical basis and inner workings; (2) A continuously learning system that attempts to predict future states and (3) Immersion in a sufficiently rich social environment from which models of yourself can be built. As Chris Frith put it, "Consciousness is for other people". Building a conscious robot, it seems, would require that we first build other conscious robots like it with whom it can then interact... unless we ourselves play that role.

Consciousness for a Social Robot Is Not Piecemeal

Juyang Weng

Dept. of Computer Science and Engineering,
Cognitive Science Program,
Neuroscience Program
Michigan State Univ.
East Landing, USA

weng@cse.msu.edu

I am glad that Janet Wiles raised the important subject of consciousness in the 11-th year of the newsletter. Consciousness appears to be behaviors arising through rich interactions among all brain functions from parallel computation of neurons. This seems to be true for all species in the animal kingdom.

The term Autonomous Mental Development (AMD)—the name of this newsletter—was originally meant not to be too restrictive. Steven Pinker 1997 wrote: "The Mind is what the brain does." When the term—Autonomous Mental Development—was coined and then discussed during 2000 for the policy forum article titled Autonomous Mental Development by Robots and Animals for Science, a respected scholar suggested "cognitive" instead of "mental". However, the subjects of autonomous development are numerous, including, e.g., body, perception, cognition, behavior, language, motivation, personality, consciousness, and intelligence. Where did we want to draw the boundary for the scope of autonomous development? Some journals and societies have focused on a relatively narrower subject. For example, a journal is called Hippocampus. However, if we understand any brain function in a piecemeal way, we probably cannot do justice to it. The same is true for any brain area. Cognition is still a piecemeal, although the subject is very rich.

Janet Wiles correctly mentioned different aspects of awareness or consciousness, such as body, social spatial relationships, individual identities, subjective world, subjective sense of self, ability to decide, and unity of a stable sensorimotor experience. Except for the first aspect (body), all the following aspects belong to the brain. However, all these aspects are very tightly related. One cannot do without other. Therefore, it seems meaningless to ask what aspects a social robot needs but not other aspects.

This is what we feel from our developmental model of a brain—Developmental Network (DN) (Weng, 2011)—which appears to be capable of developing all the above capabilities, at least in principle at this point of time. With its embodiments from WWN-1 to WWN-9, our goal for DN is to enable robots to be artificially "alive" with all major sensing modalities and motoric modalities while leaning and performing concurrently in real time. The DN has not reached this goal yet. Nevertheless, we feel that any social robot cannot claim to have one aspect of consciousness without a computational brain model that is capable of developing all.

For example, a body appears to be essential for developing a subjective sense of self as well as an objective sense of other objects in the environment.

The self can see its physical self-body that is physically separate from other individual bodies and can act differently from other individual bodies. Furthermore, the body observes physical effects when it acts on the physical environment (e.g., move a toy, push a rat, or move the self-body). Self-awareness is supported by such rich experience.

The self is relative to others. When another body moves relative to the self, neurons in the self-brain learn to segment the body's image patch on the self-retinae from the cluttered background. This is because each neuron dynamically trims its pre-synaptic connections that do not match well with the corresponding synaptic conductance (weight). In this way, it automatically cuts off leaked-in background pixels. This automatic segmentation is essential for autonomous visual learning directly from cluttered scenes. Therefore, real-time video with synchronized action stream, instead of many static images in some standard image data sets (e.g., ImageNet) no matter how many, is necessary for autonomously developing visual awareness. A child can recognize his mother in a static photo because he has mostly seen his mother moving around before he sees the static photo.

During this body sensing-and-acting process, the agent needs all other aspects that Janet Wiles mentioned: social spatial relationships (e.g., locate and recognize rat in the cluttered scene but the self is here), individual identities (e.g., toy, rat, self), subjective world (e.g., rat may attack), subjective sense of self (e.g., self-body moves forward), the ability to decide (e.g., flee or fight), and the unity of a stable sensorimotor experience (e.g., memorize and recall past spatiotemporal events with the rat). Missing one aspect implies that all other aspects are not able to autonomously develop. E.g., not locating and recognizing the rat makes it impossible to memorize the rat events.

Here we have considered all handcrafted capabilities to be too brittle to qualify for a part of consciousness. We know that all biological mental capabilities are autonomously developed. If one writes the software of iRat for an aspect of consciousness above but it must work for all possible situations, he ends up with many pieces of software that do not integrate well. The resulting robot is therefore brittle, not a stable social robot. This is the dilemma of the behavior-based automaton

handcrafted by Rodney Brooks (Brooks, 1991) because the automaton is symbolic. But the behavior-based approach championed by Brooks and others has played a positive role in robotics because roboticists were limited by 3-D monolithic representation of the environment (e.g., the 3-D map that Google autonomous cars use, the RatSLAM algorithm, and all other SLAM algorithms). All 3D maps are too weak and brittle for social robots in the real world.

Critical for integrated neuronal representations, each neuron in the brain generally has both external components in its representation—sensory and motoric—as explained by the DN model. I predict that the "place cell" work (O'Keefe & Dostrovsky, 1971) of the 2014 Nobel Award has told only a half because place is a sensory account, not a motoric (e.g., intent) account. The same is true for the "grid cells" work of the same 2014 Nobel Award.

Inside a DN such many "pieces of software" automatically and incrementally emerge. They are always integrated while DN "lives" on, because the DN learns an Emergent Turing Machine (Weng, 2015). Humans observe some aspects of DN behaviors and might interpret them as consciousness. But internally, the behaviors arise from internal firing neurons that win through parallel competition. The rule-like mechanisms of the brain can be explained in terms of Emergent Turing Machine (Weng, 2015). There is no design option to choose aspects of consciousness, because the computational model of DN is at neuronal level while DN learning is at the sensorimotor level. This is a fundamental difference between developmental models (connectionist) and the traditional symbolic agent models.

In summary, let us all pay attention to computational underpinning of consciousness—brain-scale models—so that external behavior accounts (including consciousness) are better understood in terms of precise computations.

Brooks R. (1991). Intelligence without representation, Artificial Intelligence, 47, 139-160.
O'Keefe, J., and Dostrovsky, J. (1971). The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. Brain research 34, 171-175.
Weng J. (2011). Why have we passed "neural networks do not abstract well"? Natural Intelligence: the INNS Magazine, 1(1) 13-22.
Weng J. (2015). Brain as an Emergent Finite Automaton: A Theory and Three Theorems,
International Journal of Intelligent Science, 5, 112-131.

The Darker Sides of Robot Social Awareness

Guy Hoffman
hoffman@idc.ac.il

Moran Cerf
m-cerf@kellogg.northwestern.edu

Minimal social awareness required of autonomous interactive robots can include rudimentary behaviors and traits, such as awareness of spatial relationships, individual identities, and interaction histories (Wiles, 2014). These traits likely exist in a broad range of non-human organisms.

However, when considering social robots designed specifically for human interaction, an expanded notion of social awareness should be considered. This notion includes not just basic interpersonal mechanisms, but also those that are particular to the human experience, including cultural and societal aspects of behavior and awareness. In addition, are there not only minimal requirements, but also upper limits on acceptable social awareness? Some "higher" human social behaviors are generally viewed as positive and beneficial to society at large. These include self-control and delayed gratification (Baumeister and Juola Exline, 1999), altruism (Fehr and Fischbacher, 2003), as well as perspective-taking and empathy (Underwood and Moore, 1982).

That said, behavioral research also indicates that human social behavior is marked by a number of "darker" tendencies and biases. Humans are favorable toward in-group members (Mullen et al., 1992), display racial and gender stereotypes (Fiske, 1998), engage in deceit (Gino et al., 2009); and are liable to fall into group thinking, peer-pressure, and conformity (Asch, 1951).

Given the extent to which Human-Robot Interaction relies on human behaviorial research, both to design robot behavior and to perceive and analyze human behavior, how should roboticists take these darker sides of human social awareness into account?

Should robots mimic negative patterns of social awareness to better pass in human society? Should they cater to them to be more effective in interacting, understanding, and persuading humans? Or could robots present an opportunity to nudge human behavior towards more positive social behavior?

Note that both positive and negative social biases are not necessarily irrational or non-optimal, and could be beneficial and efficient for individuals and groups (Levine, 1998). Similarly, a perfectly rational robotic agent could also be acting in a socially biased

Guy Hoffman

Media Innovation Lab,
IDC Herzliya,
Herzliya, Israel

Moran Cerf

Kellogg School
of Management,
Northwestern University
Evanston, IL, USA

manner.

Consider, for example, the case of an automatic sliding door. By some measure, this is a very simple robot interacting with humans. It has a single sensor and actuator, and makes a straightforward "decision" of opening the doorway for approaching humans. Now imagine this robot enhanced by a camera and face-recognition software and programmed to prevent the entry of recognized shoplifters. Taking this idea one step further, the store's owner could request the installation of software that prevents the entry of people who are classified by a machine learning and pattern recognition algorithm as having a high likelihood of being shoplifters, or even of just having bad credit.

Has such a discriminating autonomous door acquired some level of negative social awareness?

These questions gives rise to three possible ways for researchers in Human-Robot Interaction to address negative social awareness when designing interactive robots:

The first approach is to develop robots that take into account human negative social behaviors. These robots would be more similar to us, incorporating our negative biases, and more adept to us, taking these biases into account when modeling humans.

The second approach would suggest having robots be agnostic or neutral with respect to human social biases. Those robots will be merely functional and will neither provide nor understand social patterns. Such socially lacking robots will not suffer from biases, but may also be less successful in generating a high quality of interaction with humans.

A third approach would be to design robots that are not merely not susceptible to human negative biases, but purposefully embody positive aspects of human social behavior. These robots, personifying the "better angels" of human nature (Lahti and Weinstein, 2005), may both interact successfully with humans and also help tame our own negative social behaviors.

That is, rather than acting as proxies for our own social shortcomings, robots can be thought of as tools to support more positive social awareness based on an agreed set of rules, effectively improving on human social awareness. Instead of being bounded by the same biases that humans have a hard time shaking, such as racism, dishonesty, and conformity, researchers can design robots that specifically support values like equality, honesty, and independent thinking. Through interaction, they might shape human behavior and serve as guides for more desirable behavior.

To summarize, we ask not only about the minimum set of social awareness required to simulate consciousness in an interactive robot, but also about acceptable upper bounds, given that human social awareness often leads to negative biases and behaviors. Designing robots guided by some social responsibility may shape their interaction with humans, and in turn steer us towards acting more positively.

J. Wiles, "Will Social Robots Need to Be Consciously Aware?" IEEE CIS Newsletter of the Autonomous Mental Development Technical Committee, vol. 11, no. 2, pp. 14–15, 2014.
R. F. Baumeister and J. Juola Exline, "Virtue, Personality, and Social Relations: Self-Control as the Moral Muscle," Journal of personality, vol. 67, no. 6, pp. 1165–1194, 1999.
E. Fehr and U. Fischbacher, "The nature of human altruism," Nature, vol. 425, no. 6960, pp. 785–791, 2003.
B. Underwood and B. Moore, "Perspective-taking and altruism." Psychological Bulletin, vol. 91, no. 1, p. 143, 1982.
B. Mullen, R. Brown, and C. Smith, "Ingroup bias as a function of salience, relevance, and status: An integration," European Journal of Social Psychology, vol. 22, no. 2, pp. 103–122, 1992.
S. Fiske, "Stereotyping, prejudice, and discrimination," The handbook of social psychology, p. 357, 1998.
F. Gino, S. Ayal, and D. Ariely, "Contagion and differentiation in unethical behavior: the effect of one bad apple on the barrel." Psychological science, vol. 20, no. 3, pp. 393–8, Mar. 2009.
S. E. Asch, "Effects of group pressure upon the modification and distortion of judgments," Groups, leadership, and men. S, pp. 222–236, 1951.
D. K. Levine, "Modeling altruism and spitefulness in experiments," Review of economic dynamics, vol. 1, no. 3, pp. 593–622, 1998.
D. C. Lahti and B. S. Weinstein, "The better angels of our nature: group stability and the evolution of moral tension," Evolution and Human Behavior, vol. 26, no. 1, pp. 47–63, 2005.

Designing robots for social awareness

Janet Wiles

Complex & Intelligent Systems Research Group, School of Information Technology & Electrical Engineering,
The University of Queensland
St Lucia. Australia

j.wiles@uq.edu.au

I thank the commentators for their interesting and diverse perspectives. The initial dialog raised the question of the types of awareness required to underpin social interaction, listing as a start an awareness of self, others and social spatial relationships. The commentaries point out that these are a first step but are not sufficient. They call for reframing the questions, drilling deeper into the competencies required, and expanding the notion of social awareness, drawing on extensive resources in existing social robotics research. Given the breadth of the responses, the response below takes a couple of points from each that bear on design issues.

Social sub-systems and their differential responses to humans and robots

Kitano (Kitano, 2015) distinguishes between a robot's intrinsic consciousness and human perception (and hence attribution) of consciousness. She invokes the non-conscious social machinery of human brains, citing recent neuroimaging studies by Takahashi and colleagues (Takahashi et al., 2014) who found two components in the impressions of opponents in a competitive game: first, participants attributed mental function to humans much more strongly than computers (which they called mind-holderness); and second, participants' "behavioural entropy" was equal for humans and computers (which they called mind-readerness). Activity in different systems of brain regions was correlated with these two components.

Kitano's commentary draws attention to important distinctions between a person's attributions as an observer, which can make use of the default social mind; and first-person conscious experience, which then further divides into qualia (non-reportable) and reflective (or reportable) cognitive and perceptual states. This work alerts the social robot designer to the fact that several subsystems are active in social interactions, and moreover, they are differentially active in human-robot interaction.

Design requirements for awareness

To engineer a companion robot, Weber and Wermter (W&W) (Weber and Wermter, 2015) point out that the goal of reliable, safe and meaningful interaction with people will drive social robots towards apparent awareness. For such an engineering goal, they point out that whether the robot itself is fully self-aware may remain unknown and is possibly irrelevant. Takahashi's component "mind-readerness" meshes neatly with this view.

W&W's discussion of engineering requirements is illuminative, drawing on demonstrations of recently developed social robots: First, they identify precursors to communication—assessing an interlocutor's field of view, their readiness to communicate and whether they share attention to an object. They then consider the design of the cognitive system, and suggest that the ingenious representations afforded by mirror neurons can serve as a powerful method for common referencing between observation and action, mind-reading, and learning from imitation. Since their first discovery, the design challenge has always been to understand—even in principle—how a mirror neuron could determine which observations and actions should share a reference. Wermter et al's MirrorBot (Wermter et al., 2005) provides one approach. Following the biological inspiration of mirror neurons, W&W point out that methods for verbal communication can draw on human communication strategies, including emotional state and intentions. They also add that a robot should display traits of consciousness (although for an engineering solution, it would be useful to unpack such traits further). They then note that verbal communication entails user expectations, and suggest that robots could use speech or gesture to convey their internal states, with algorithmic analogues of feelings to increase interpretability. They conclude with a list of requirements for self-awareness: alertness, structured neural representations, attention selection, short-term memory, motivation and emotions, providing a useful set of design requirements for engineering social robots.

Developmental modelling of the self: the bio-social approach

The ease with which a normally developing child acquires social competence belies the complexity of the underlying system. Hart and Scassellati (H&S) (Hart and Scassellati, 2015) point out that a diverse set of developmentally emerging capabilities are precursors to proficient social behaviour, including an understanding of body and senses; and reasoning about mental states. H&S stress that social competency is not monolithic, but encompasses a range of competencies, and that designing a robot for social interaction requires determining which of these social skills are desired. They illustrate the task of robot self-modelling by developing understanding from the robot's own operation (Hart and Scassellati, 2015). Their robot uses its body and sensors in conjunction with each other: using its vision system to develop a kinematic model, then using the kinematic system for further refinement of

its self-modelling and also to reason about its environment, including the challenging task of inferring positions from mirror images. A robot that is aware that its own mental state differs from others can reflect on self to reason about others and vice versa, social skills that H&S indicate play roles in learning, collaboration, and competition.

H&S predict that self-awareness will become increasingly important to robot interactions in future. Their path to such a future is based on knowledge from psychology and neuroscience, building on their current robot's development of self-models. As their work has shown, child development provides existence proofs as well as useful guidance for developmental sequences that lead to effective social behaviour. Robotics has the potential to play an interesting role in exploring these ideas and beyond, testing whether social competence requires that robots recapitulate the developmental paths of children or whether there are many paths to engineering social robots.

Requirements for unscripted interactions

On the broader question of what it would take to build a consciously aware robot, Cleeremans (Cleeremans, 2015) points out that there are theories, but no consensus to use as a guide. He identifies key missing aspects as the basic components of wants, desires, fears and hopes, but also intentions. Robots are particularly helpful in thinking about consciousness, because of the engineering requirement for a working prototype, which is absent in thought experiments. What would it take to build a conscious robot? Cleeremans uses the very attempt to show up the limitations in current social robots.

Like H&S, Cleeremans' commentary addresses the modelling of mental states of self and others, however, he emphasises the importance of unscripted interactions, beyond the many daily scripted ones, and the critical role of mental-state modelling. He identifies features that make unscripted interactions possible as the representation of the mental states and intentions of other agents; computational resources for simulation; a predictive system that is continuously learning; and a rich social environment with the necessary experiences. The mental states of self and other are not available to inspection and building a model of how agents work is an inverse problem with a many-to-many mapping. Cleeremans puts forward a "provocative claim" (close to his own view?) that Theory of Mind comes first: Learning to anticipate consequences leads to models of others' mental states, and such models then lead to self-insight.

Designing better angels for robot natures

Hoffman and Cerf (2015) (H&C) (Hoffman and Cerf, 2015) point out that beyond basic mechanisms, there are cultural and societal aspects of awareness which should be considered. Perfect rationality could underpin social biases that may be beneficial for individuals and groups, aiming for successful interactions, yet also embody – albeit unintentionally – the "darker" tendencies of human nature, including racism, stereotyping, group thinking, etc.

H&C's commentary is a reminder that design principles for competency are not enough: ethics is intrinsic to social interactions. If they are not explicitly considered, the emergent behaviours of rational systems may unwittingly result in socially negative outcomes. H&C raise the possibility of designing robots that personify the "better angels" of human nature and that are tools for positive social awareness. This is an important and timely call. The challenging question for social robotics is how to engineer such better angels for social robots. H&C suggest an approach based on an agreed set of rules to support values such as equality, honesty and independent thinking.

Given the important distinction between underlying mechanisms and emergent outcomes, it is not possible to simply specify a desired outcome (such as equality) and expect a robot designer to know what mechanisms will achieve it. If social robotics as a research field is to act on H&C's call, research programs will need to propose sets of rules for positive social outcomes, mechanisms to engineer them, and empirical studies that monitor progress towards their realisation.

Integrative systems and specialist components

Weng (Weng, 2015) takes a different position from the other commentaries, starting from the assertion that all aspects of awareness are tightly interrelated and each cannot function without the others. He calls attention to his brain scale models which maintain integration while enabling automatic incremental emergence, and argues that these are the way to understand the precise computations underpinning consciousness. Weng's commentary differs from the others on the question of whether awareness is monolithic. It can be rephrased as the question of whether awareness is an X-complete problem, a class of problems in which solution to one problem in the class provides (or requires) a solution to all, such as NP-complete (Garey and Johnson, 1979) or AI-complete (Kirsh, 1986).

Integrative systems are needed in modelling, but we should be sceptical of approaches

that exclude progress on understanding the biological sub-systems of different neural regions. Weng raised the example of the hippocampus as understanding a brain function in "a piecemeal way". Brains do have differentiated neural circuits and study of the anatomy and electrophysiology of these circuits has provided many insights into spatial awareness (O'Keefe and Dostrovsky, 1971; Hafting et al., 2005); computational modelling of neural circuits in the hippocampus has led to insights into the role of neurogenesis in spatial memory (Aimone et al., 2009) and grid cells in navigation (Milford et al., 2010); and from an engineering perspective, bio-mimicry of place and grid cells has led to practical mapping algorithms, which have been demonstrated on the large scale real-world task of mapping the entire 66km of a suburb (Milford and Wyeth, 2008). At a theoretical level, the addition of a SLAM module provides an environment-centric map, which is vastly superior to idiopathic path integration in noisy systems (Vickerstaff and Cheung, 2010). Each neural circuit in the sub-regions of the hippocampus and para-hippocampal areas has a useful functionality. Models of single neural regions may be brittle, but together they are flexible and powerful. An integrative system alone without such specialised neural circuits has yet to demonstrate the impressive navigational achievements and concomitant spatial awareness that detailed bio-mimicry of these circuits has achieved.

Social robotics needs an understanding of both integrative systems and the diverse set of specialist skills contributed by many different sub-systems. It's the combination and collaboration between the two that leads to scientific understanding and effective engineering design.

Y. Kitano, "For Whom Robots are Conscious?," IEEE CIS Newsletter on Autonomous Mental Development, vol. 12 (1) (this volume), 2015.
H. Takahashi, K. Terada, T. Morita, S. Suzuki, T. Haji, H. Kozima, M. Yoshikawa, Y. Matsumoto, T. Omori, and M. Asada, "Different impressions of other agents obtained through social interaction uniquely modulate dorsal and ventral pathway activities in the social human brain," Cortex, vol. 58, pp. 289-300, 2014.
C. Weber and S. Wermter, "No arguments against consciously aware robots," IEEE CIS Newsletter on Autonomous Mental Development, vol. 12 (1) (this volume), 2015.
S. Wermter, C. Weber, M. Elshaw, V. Gallese, and F. Pulvermüller, "Grounding neural robot language in action. Biomimetic Neural Learning for Intelligent Robots," in Intelligent Systems, Cognitive Robotics, and Neuroscience, 2005.
J. W. Hart and B. Scassellati, "Self-awareness and social competencies," IEEE CIS Newsletter on Autonomous Mental Development, vol. 12 (1) (this volume), 2015.
J. W. Hart and B. Scassellati, "Robotic Self-Models Inspired by Human Development," in Metacognition for Robust Social Systems, 2010.
A. Cleeremans, "Commentary on robot consciousness," IEEE CIS Newsletter on Autonomous Mental Development, vol. 12 (1) (this volume), 2015.
G. Hoffman and M. Cerf, "The darker side of robot social awareness," IEEE CIS Newsletter on Autonomous Mental Development, vol. 12 (1) (this volume), 2015.
J. Weng, "Consciousness for a Social Robot Is Not Piecemeal," IEEE CIS Newsletter on Autonomous Mental Development, vol. 12 (1) (this volume), 2015.
M. R. Garey and D. S. Johnson, "Computers and intractability: a guide to NP-completeness," ed: WH Freeman New York, 1979.
D. Kirsh, "The term "AI-complete" is attrib to Fanya Montalvo in "Second-generation AI theories of learning"," Behavioral and Brain Sciences, vol. 9, pp. 658-659, 1986.
J. O'Keefe and J. Dostrovsky, "The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat," Brain Research, vol. 34, pp. 171-175, 1971.
T. Hafting, M. Fyhn, S. Molden, M.-B. Moser, and E. I. Moser, "Microstructure of a spatial map in the entorhinal cortex," Nature, vol. 436, pp. 801-806, 2005.
J. B. Aimone, J. Wiles, and F. H. Gage, "Computational influence of adult neurogenesis on memory encoding," Neuron, vol. 61, pp. 187-202, 2009.
M. J. Milford, J. Wiles, and G. F. Wyeth, "Solving navigational uncertainty using grid cells on robots," PLoS computational biology, vol. 6, p. e1000995, 2010.
M. J. Milford and G. F. Wyeth, "Mapping a suburb with a single camera using a biologically inspired SLAM system," Robotics, IEEE Transactions on, vol. 24, pp. 1038-1053, 2008.
R. J. Vickerstaff and A. Cheung, "Which coordinate system for modelling path integration?," Journal of theoretical biology, vol. 263, pp. 242-261, 2010.

New Dialogue Initiation

Representational redescription: the next challenge?

Stéphane Doncieux

ISIR,
Université Pierre et Marie Curie,
Paris, France
doncieux@isir.upmc.fr

Drawing inspiration from developmental psychology, it has been suggested to build cognitive architectures that allow robots to progressively acquire abstract representations (Guerin et al., 2013). Humans don't have a single optimal representation of the problems they solve. They can redescribe the information they have acquired in different formats (Karmiloff-Smith, 1995). It allows them to explore different representations and use multiple problem solving strategies, from lowlevel systematic search to abstract reasoning (Evans, 2003).

Representational redescription is the ability to change the way information is stored and manipulated, to make further treatments easier and more efficient. A representation is the description of some data in a given format. The lowest level possible for formats is the raw format of sensors and effectors. Some examples of high level representation can be drawn from artificial intelligence and machine learning communities: markov decision processes formalism, first order logic or neural networks. Changing the representation allows usage of different problem solving strategies. Adapted representations make computations easier by relying on a small set of relevant primitives instead of a big set of unstructured data.

Use a single representation or change representations over time?

Humans may use representational redescription because of physiological constraints. The genome contains twenty thousand genes to describe the whole body, including the brain with its hundred billions of neurons. Such a small number of genes may not be enough for a genetic transmission of sophisticated representations. Does it necessarily mean that robots should follow the same path? Human representational redescription may also be an advantage rewarded by evolutionary pressure because of the adaptation ability it has resulted in. Would it help robots to face open environments? This would undoubtedly be an interesting feature. In the following, we will consider the questions that it raises.

Where to start?

Sensorimotor data first need to be observed before they can be redescribed in a format that allows an agent to better understand what happened and eventually to reproduce it. Babies have grasping or sucking reflexes that allow them to start interacting with surrounding objects before they can perform more complex actions. Guerin et al. suggest using a similar set of innate sensorimotor schemas to bootstrap the process (Guerin et al., 2013). How to choose this set of primitive schemas and where to stop? If we, as roboticists, do know how to implement an efficient grasping behavior, why should we start with an inefficient grasping reflex? A sophisticated grasping behavior may allow the robot to rapidly and efficiently interact with objects, thus generating a lot of useful data to learn about them. Where should we then put the frontier between the schemas that are provided to the robot and the ones that should be discovered? Providing efficient behaviors is clearly a convenient way to bootstrap the process. Are there other alternatives?

Evolution shaped development, but could it be also involved in the representational redescription process?

Evolution has shaped, over millions of years, living beings and their development process. But beyond this first evo-devo relation, evolutionary mechanisms may also be at play during development and learning. The principles of variation and selection have contributed to the success of evolutionary computation because of their simplicity, robustness and versatility. They have been used in a robotics context for more than twenty years (Doncieux et al., 2015), and were notably able to generate non trivial behaviors with neural networks. They are also believed to be the primary mechanisms in development, both for learning motor schemas and for selecting problem solving strategies (Guerin et al., 2013). They could then have a significant role to play in the representational redescription process, in particular thanks to their ability to generate controllers relying on the most simple representations, i.e. sensorimotor data. Furthermore, this hypothesis may be biologically plausible, as evolutionary principles can be implemented along with neural mechanisms (Fernando et al., 2012). Evolution could then be involved in brain functions and thus in development and learning.
Should representation formats be given a priori or should it emerge from the developmental process?

Representational redescription requires the availability of the representation formats in which the redescription is expected to occur. A first possibility would be to provide the

agent with different representation formats, like first order logic or markov decision processes formalism, for instance. Dedicated machine learning algorithms could extract them from a lower level representation, e.g. the sensorimotor flow. An alternative would be to use a versatile connectionist formalism and rely on deep learning algorithms to redescribe lower layers representations to more abstract ones. The first alternative is a somewhat top-down approach in which learning and decision algorithms are available from the very beginning. The developmental process "just" needs to represent sensorimotor data in the corresponding format for the system to exhibit high level cognitive abilities. The second is a bottom-up approach in which higher level representations emerge progressively and where the corresponding problem solving strategies will also need to emerge.

Does provided knowledge limit developmental abilities?

Providing knowledge allows one to take shortcuts in the developmental process: no need to discover what is provided and the corresponding developmental time is then saved. Providing sensorimotor schemas or representation formats constrains what the agent can do, what it will observe and what it will extract from these observations. If the agent is expected to face an open environment, isn't it a limit to its adaptive abilities? Are there conflicts between shortening developmental time and having an open-ended developmental process?

How to make a robot endowed with representational redescription transparent?

Giving a robot the ability to change its representations and problem solving strategies may make it difficult to understand for a human. A non-expert may have trouble predicting what the system will actually do and what it understands from its environment. Making such robots transparent may then be critical for them to be used in practice, in particular if they are to enter our everyday environment. How could it be achieved?

Doncieux, S., Bredeche, N., Mouret, J.-B., and Eiben, A. (2015). Evolutionary robotics: what, why, and where to. Frontiers in Robotics and AI, 2.
Evans, J. (2003). In two minds: dual-process accounts of reasoning. Trends in Cognitive Sciences, 7(10):454–459.
Fernando, C., Szathmary, E., and Husbands, P. (2012). Selectionist and evolutionary approaches to brain function: a critical appraisal. Frontiers in Computational Neuroscience, 6(April):1–28.
Guerin, F., Kruger, N., and Kraft, D. (2013). A Survey of the Ontogeny of Tool Use : from Sensorimotor Experience to Planning. IEEE Transactions on Autonomous Mental Development, 5(1):18–45.
Karmiloff-Smith, A. (1995). Beyond Modularity: A Developmental Perspective on Cognitive Science. The MIT Press.

IEEE TAMD Table of Contents

Volume 7, Issue 1, March 2015

Editorial IEEE Transactions on Autonomous Mental Development

A. Cangelosi

Ecological Active Vision: Four Bioinspired Principles to Integrate Bottom–Up and Adaptive Top–Down Attention Tested With a Simple Camera-Arm Robot

D. Ognibene, G. Baldassare

Vision gives primates a wealth of information useful to manipulate the environment, but at the same time it can easily overwhelm their computational resources. Active vision is a key solution found by nature to solve this problem: a limited fovea actively displaced in space to collect only relevant information. Here we highlight that in ecological conditions this solution encounters four problems: 1) the agent needs to learn where to look based on its goals; 2) manipulation causes learning feedback in areas of space possibly outside the attention focus; 3) good visual actions are needed to guide manipulation actions, but only these can generate learning feedback; and 4) a limited fovea causes aliasing problems. We then propose a computational architecture ("BITPIC") to overcome the four problems, integrating four bioinspired key ingredients: 1) reinforcement-learning fovea-based top-down attention; 2) a strong vision-manipulation coupling; 3) bottom-up periphery-based attention; and 4) a novel action-oriented memory. The system is tested with a simple simulated camera-arm robot solving a class of search-and-reach tasks involving color-blob "objects." The results show that the architecture solves the problems, and hence the tasks, very efficiently, and highlight how the architecture principles can contribute to a full exploitation of the advantages of active vision in ecological conditions.

A Simplified Cerebellar Model with Priority-based Delayed Eligibility Trace Learning for Motor Control

V.A. Shim, C. S. N. Ranjit, Bo Tian, Miaolong Yuan, Huajin Tang

The study of cerebellum has resulted in a common agreement that it is implicated in motor learning for movement coordination. Learning governed by error signal through synaptic eligibility traces has been proposed to be a learning mechanism in cerebellum. In this paper, we extend this idea and suggest a simplified and improved cerebellar model with priority-based delayed eligibility trace learning rule (S-CDE) that enables a mobile robot to freely and smoothly navigate in an environment. S-CDE is constructed in a brain-based device which mimics the anatomy, physiology, and dynamics of cerebellum. The input signal in terms of depth information generated from a simulated laser sensor is encoded as neuronal region activity for velocity and turn rate control. A priority-based delayed eligibility trace learning rule is proposed to maximize the usage of input signals for learning in synapses on Purkinje cell and cells in the deep cerebellar nuclei of cerebellum. Error signal generation and input signal conversion algorithms for turn rate and velocity are designed to facilitate training in an environment containing turns of varying curvatures. S-CDE is tested on a simulated mobile robot which had to randomly navigate maps of Singapore and Hong Kong expressways.

Mental States, EEG Manifestations, and Mentally Emulated Digital Circuits for Brain-Robot Interaction

S. Bozinovski, A. Bozinovski

This paper focuses on electroencephalogram (EEG) manifestations of mental states and actions, emulation of control and communication structures using EEG manifestations, and their application in brain-robot interactions. The paper introduces a mentally emulated demultiplexer, a device which uses mental actions to demultiplex a single EEG channel into multiple digital commands. The presented device is applicable in controlling several objects through a single EEG channel. The experimental proof of the concept is given by an obstacle-containing trajectory which should be negotiated by a robotic arm with two degrees of freedom, controlled by mental states of a human brain using a single EEG channel. The work is presented in the framework of Human-Robot interaction (HRI), specifically in the framework of brain-robot interaction (BRI). This work is a continuation of a previous work on developing mentally emulated digital devices, such as a

mental action switch, and a mental states flip-flop.

Can Real-Time, Adaptive Human–Robot Motor Coordination Improve Humans' Overall Perception of a Robot?

Qiming Shen, K. Dautenhahn, J. Saunders, H. Kose

Previous research on social interaction among humans suggested that interpersonal motor coordination can help to establish social rapport. Our research addresses the question of whether, in a human-humanoid interaction experiment, the human's overall perception of a robot can be improved by realizing motor coordination behavior that allows the robot to adapt in real-time to a person's behavior. A synchrony detection method using information distance was adopted to realize the real-time human-robot motor coordination behavior, which guided the humanoid robot to coordinate its movements to a human by measuring the behavior synchrony between the robot and the human. The feedback of the participants indicated that most of the participants preferred to interact with the humanoid robot with the adaptive motor coordination capability. The results of this proof-of-concept study suggest that the motor coordination mechanism improved humans' overall perception of the humanoid robot. Together with our previous findings, namely that humans actively coordinate their behaviors to a humanoid robot's behaviors, this study further supports the hypothesis that bidirectional motor coordination could be a valid approach to facilitate adaptive human-humanoid interaction.

Volume 7, Issue 2, June 2015

Sparsity-Constrained fMRI Decoding of Visual Saliency in Naturalistic Video Streams

X. Hu, C. Lv, G. Cheng, J. Lv, L. Guo, J. Han, T. Liu

Naturalistic stimuli such as video watching have been increasingly used in functional magnetic resonance imaging (fMRI)-based brain encoding and decoding studies since they can provide real and dynamic information that the human brain has to process in everyday life. In this paper, we propose a sparsity-constrained decoding model to explore whether bottom-up visual saliency in continuous video streams can be effectively decoded by brain activity recorded by fMRI, and to examine whether sparsity constraints can improve visual saliency decoding. Specifically, we use a biologically-plausible computational model to quantify the visual saliency in video streams, and adopt a sparse representation algorithm to learn the atomic fMRI signal dictionaries that are representative of the patterns of whole-brain fMRI signals. Sparse representation also links the learned atomic dictionary with the quantified video saliency. Experimental results show that the temporal visual saliency in video stream can be well decoded and the sparse constraints can improve the performance of fMRI decoding models.

Motor-Primed Visual Attention for Humanoid Robots

L. Lukic, A. Billard, J. Santos-Victor

We present a novel, biologically inspired, approach to an efficient allocation of visual resources for humanoid robots in a form of a motor-primed visual attentional landscape. The attentional landscape is a more general, dynamic and a more complex concept of an arrangement of spatial attention than the popular "attentional spotlight" or "zoom-lens" models of attention. Motor-priming of attention is a mechanism for prioritizing visual processing to motor-relevant parts of the visual field, in contrast to other, motor-irrelevant, parts. In particular, we present two techniques for constructing a visual "attentional landscape". The first, more general, technique, is to devote visual attention to the reachable space of a robot (peripersonal space-primed attention). The second, more specialized, technique is to allocate visual attention with respect to motor plans of the robot (motor plans-primed attention). Hence, in our model, visual attention is not exclusively defined in terms of visual saliency in color, texture or intensity cues, it is rather modulated by motor information. This computational model is inspired by recent findings in visual neuroscience and psychology. In addition to two approaches to constructing the attentional landscape, we present two methods for using the attentional landscape for driving visual processing. We show that motor-priming of visual attention can be used to very efficiently distribute limited computational resources devoted to the visual processing. The proposed model is validated in a series of experiments conducted with the iCub robot, both using the simulator and the real robot.

A Probabilistic Concept Web on a Humanoid Robot

H. Celikkanat, G. Orhan, S. Kalkan

It is now widely accepted that concepts and conceptualization are key elements towards achieving cognition on a humanoid robot. An important problem on this path is the grounded representation of individual concepts and the relationships between them. In this article, we propose a probabilistic method based on Markov Random Fields to model a concept web on a humanoid robot where individual concepts and the relations between them are captured. In this web, each individual concept is represented using a prototype-based conceptualization method that we proposed in our earlier work. Relations between concepts are linked to the cooccurrences of concepts in interactions. By conveying input from perception, action, and language, the concept web forms rich, structured, grounded information about objects, their affordances, words, etc. We demonstrate that, given an interaction, a word, or the perceptual information from an object, the corresponding concepts in the web are activated, much the same way as they are in humans. Moreover, we show that the robot can use these activations in its concept web for several tasks to disambiguate its understanding of the scene.

Action Priors for Learning Domain Invariances

B. Rosman, S. Ramamoorthy

An agent tasked with solving a number of different decision making problems in similar environments has an opportunity to learn over a longer timescale than each individual task. Through examining solutions to different tasks, it can uncover behavioral invariances in the domain, by identifying actions to be prioritized in local contexts, invariant to task details. This information has the effect of greatly increasing the speed of solving new problems. We formalise this notion as action priors, defined as distributions over the action space, conditioned on environment state, and show how these can be learnt from a set of value functions. We apply action priors in the setting of reinforcement learning, to bias action selection during exploration. Aggressive use of action priors performs context based pruning of the available actions, thus reducing the complexity of lookahead during search. We additionally define action priors over observation features, rather than states, which provides further flexibility and generalizability, with the additional benefit of enabling feature selection. Action priors are demonstrated in experiments in a simulated factory environment and a large random graph domain, and show significant speed ups in learning new tasks. Furthermore, we argue that this mechanism is cognitively plausible, and is compatible with findings from cognitive psychology.

Staged Development of Robot Skills: Behavior Formation, Affordance Learning and Imitation with Motionese

E. Ugur, Y. Nagai, E. Sahin, E. Oztop

Inspired by infant development, we propose a three staged developmental framework for an anthropomorphic robot manipulator. In the first stage, the robot is initialized with a basic reach-and-denclose-on-contact movement capability, and discovers a set of behavior primitives by exploring its movement parameter space. In the next stage, the robot exercises the discovered behaviors on different objects, and learns the caused effects; effectively building a library of affordances and associated predictors. Finally, in the third stage, the learned structures and predictors are used to bootstrap complex imitation and action learning with the help of a cooperative tutor. The main contribution of this paper is the realization of an integrated developmental system where the structures emerging from the sensorimotor experience of an interacting real robot are used as the sole building blocks of the subsequent stages that generate increasingly more complex cognitive capabilities. The proposed framework includes a number of common features with infant sensorimotor development. Furthermore, the findings obtained from the self-exploration and motionese guided human-robot interaction experiments allow us to reason about the underlying mechanisms of simple-to-complex sensorimotor skill progression in human infants.

Structural Bootstrapping—A Novel, Generative Mechanism for Faster and More Efficient Acquisition of Action-Knowledge

F. Worgotter, C. Geib, M. Tamosiunaite, E.E. Aksoy

Humans, but also robots, learn to improve their behavior. Without existing knowledge, learning either needs to be explorative and, thus, slow or–to be more efficient–it needs to rely on

supervision, which may not always be available. However, once some knowledge base exists an agent can make use of it to improve learning efficiency and speed. This happens for our children at the age of around three when they very quickly begin to assimilate new information by making guided guesses how this fits to their prior knowledge. This is a very efficient generative learning mechanism in the sense that the existing knowledge is generalized into as-yet unexplored, novel domains. So far generative learning has not been employed for robots and robot learning remains to be a slow and tedious process. The goal of the current study is to devise for the first time a general framework for a generative process that will improve learning and which can be applied at all different levels of the robot's cognitive architecture. To this end, we introduce the concept of structural bootstrapping–borrowed and modified from child language acquisition–to define a probabilistic process that uses existing knowledge together with new observations to supplement our robot's data-base with missing information about planning-, object-, as well as, action-relevant entities. In a kitchen scenario, we use the example of making batter by pouring and mixing two components and show that the agent can efficiently acquire new knowledge about planning operators, objects as well as required motor pattern for stirring by structural bootstrapping. Some benchmarks are shown, too, that demonstrate how structural bootstrapping improves performance.

EDITOR **Pierre-Yves Oudeyer,** Inria and Ensta ParisTech, France, pierre-yves.oudeyer@inria.fr
EDITORIAL ASSISTANT **Fabien Benureau,** Inria and Ensta ParisTech, France
ISSN 1550-1914